

Award Number: **W81XWH-09-1-0546**

TITLE:

Oncogenicity and Selective Inhibition of ERG Splicing Variants in Prostate Cancer

PRINCIPAL INVESTIGATOR:

LUCA CARTEGNI, PhD

CONTRACTING ORGANIZATION:

**Sloan Kettering Institute for Cancer Research
New York, NY 10021**

REPORT DATE: **September 2010**

TYPE OF REPORT:

Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: (Check one)

☒ Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE 30-SEP-2010		2. REPORT TYPE ANNUAL		3. DATES COVERED (From - To) 01 SEP 2009 - 31 AUG 2010	
4. TITLE AND SUBTITLE Oncogenicity and Selective Inhibition of ERG Splicing Variants in Prostate Cancer Cancer				5a. CONTRACT NUMBER W81XWH_09-1-0546	
				5b. GRANT NUMBER P	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Luca Cartegni cartegni@mskcc.org				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Sloan Kettering Institute New York, for Cancer Research 10065				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Ft. Detrick, MD 21702				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Genetic rearrangements that result in the formation of oncogenic fusion genes are well-characterized key events in cancer development in several rare haematological malignancies and sarcomas. Recently they have been suggested to play a major role also in common solid tumors, namely in a majority of prostate cancers. ERG, a member of the ETS transcription factor family known to be involved in different oncogenic fusions, is frequently overexpressed in clinical prostate cancer, driven by its fusion to the androgen-responsive TMPRSS2 gene. In particular, The fusion of exon 1 of TMPRSS2 with exon 4 of ERG has been observed in about 50% of human prostate cases, suggesting a causative role in the development of carcinomas. Additional TMPRSS2:ERG fusions (involving different recombination sites) have been described, but their specific functions and oncogenic activities are still unclear. We have started to systematically address the complex gene expression patterns of both the endogenous ERG genes, and found that it alone can generate over 50 variants, some of which appear to correlate with tumor aggressiveness. The systematic characterization of both ERG and TMPRSS2:ERG main variants underscores differences in their biological roles and oncogenic potential, which depends on the presence of the intact Ets domain and on the configuration of the 5' UTR region. In additon, the fusion-derived variants can be preferentially downregulated by antisense-based compounds designed to specifically block their translation initiation site					
15. SUBJECT TERMS tmprss2, ERG, prostate cancer, alternative splicing, translational control, antisense					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 25	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

Table of Contents

Page

Introduction..... 3

Body..... 5

Key Research Accomplishments..... 17

Reportable Outcomes..... 17

Conclusion..... 18

References..... 20

Appendices..... 22

Supporting data..... 22

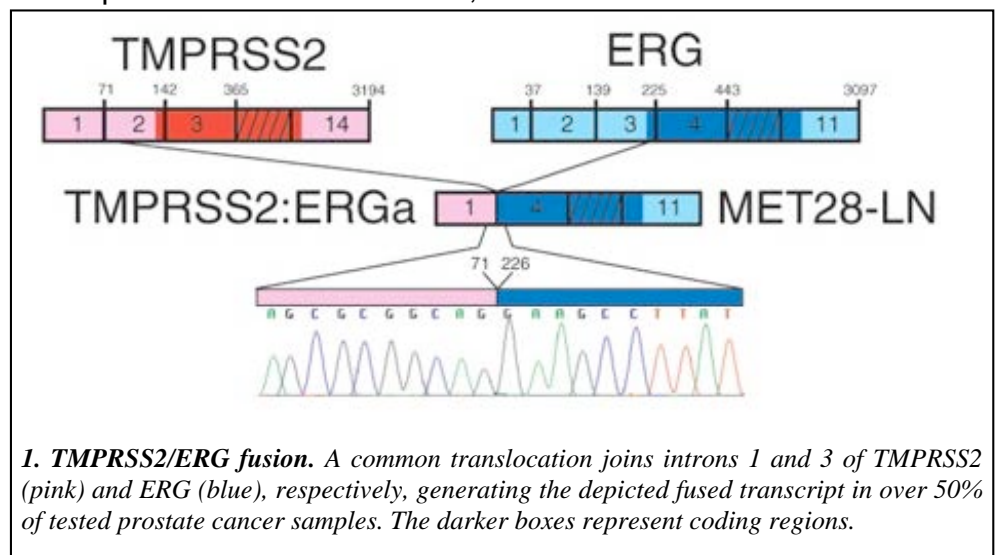
INTRODUCTION

Chromosomal translocations are key events in the development of cancer and have been very well established in haematological malignancies and sarcomas [1]. Until recently, such rearrangements were not considered to play a major role in carcinomas of epithelial origin. However, the prostate-specific identification of frequent recurrent translocations between the androgen-responsive TMPRSS2 gene and members of the Ets family of transcription factors, has forced to revisit this notion and has changed the panorama of prostate cancer biology [2]. As Ets proteins are already implicated in other oncogenic translocations, the likely consequence of their fusion to androgen responsive genes would be the acquisition of the tumorigenic properties associated to the Ets transcription factors by cells sensitive to androgen stimulation, such as prostate cells.

Prostate cancer is the most common malignancy in men in developed countries, and the leading cause of cancer-related death in males [3]. More than 80% of prostate cancers harbor fusions which typically involve the 5' region of the androgen-responsive TMPRSS2 locus (including its promoter) joined to the 3' region of various Ets genes (lacking the promoters but including all or most of the Open Reading Frame, ORF) [4]. ETV1 and ERG were the first identified 3' fusion partners of the TMPRSS2 gene [2, 5], but subsequent analysis lead to the description of additional Ets family members as 3' partners for TMPRSS2 and of other 5' partners for Ets genes[5-7]. The most common rearrangement, TMPRSS2-exon1:ERG-exon4 (T1:E4, or variant III) arises in ~50% of prostate cancer cases[4] as a fusion that joins the TMPRSS2 5' UTR to most of ERG ORF (Figure 1).

Multiple observations suggest that Ets gene fusions may play a role in the transition from prostatic intraepithelial neoplasia (PIN) to adenocarcinoma and invasion, and are associated to aggressive lesions and poor prognosis[8-12]. Overexpression of ERG in prostate cell lines activates cell invasion programs and results in the development of PINs in mice, but it is not sufficient to drive carcinogenesis [6, 9, 11, 13]. Cooperation with separate genetic lesions, such as for example pTEN loss, that dysregulate cellular proliferation and other control mechanisms is needed to trigger progression to advanced disease [8, 14, 15].

Multiple variants of the



normal ERG gene product have been described, arising from a combination of alternative splicing, polyadenylation and transcriptional initiation[16-19]. The encoded ERG protein isoforms interact with the AP1 complex to activate transcription and their activity is modulated by homo- and heterodimeric interactions among ERG and other Ets variants[20, 21]. Variability in the coding region can influence ERG activity of the TMPRSS2:ERG fusions as well, and the presence of a variant including a 72 nt alternative exon shows enhanced biological activity, especially when expressed together with other isoforms [21].

The specifics of the genomic rearrangements also introduce considerable structural heterogeneity in the 5' region. In addition to the common T1:E4 transcript, TMPRSS2 exons 1, 2 and 3 can be combined with ERG exons 2, 3, 4, 5 and 6 in various alternative splicing patterns that can generate at least 17 distinct TMPRSS2:ERG transcripts [2, 22-24] which can serve as markers for disease progression, and are correlated to the aggressiveness of the tumors [9, 10, 22]. In particular, the presence of the T2:E4 variant (TMPRSS2-exon2:ERG-exon4 or variant VI), in which the native ATG in TMPRSS2 exon 2 is in frame with the ERG ORF, is associated with pathological and clinical aspects of aggressive disease [22]. The mechanistic basis for these differences in oncogenic potential of the fusion isoforms remains to be elucidated, and it could be related to intrinsic differences in the N-terminal regions or to the effect(s) that variation in the 5' region of the mRNA can have on RNA stability or expression.

To evaluate and characterize the various ERG and TMPRSS2:ERG isoforms, we sought to assess differences between the use of 3 alternative promoters, 2 common alternative splicing events and 3 polyadenylation sites in normal tissues relative to TMPRSS2:ERG-expressing prostate tumors. These independently regulated events combine to generate 30 'main' native isoforms, some of which are also highly overexpressed in tumors. The characterization of the translation initiation sites used by the most common native ERG and fusion variants reveals the specific organization of the 5' UTR region as one of the principal determinants of their biological activity and identifies an ATG in exon 4 as a promising target for antisense-based translation inhibition in prostate cancer.

The development of specific translation-blocking compounds that can effectively and selectively reduce the levels of aberrant ERG isoforms would introduce an important set of tools to enhance our understanding of a pathway that is improperly activated in the majority of prostate cancer occurrences, and could form the basis for novel approach in their treatment.

BODY

Significant progress has been achieved in the past year on the various Aims proposed in the original proposal as detailed below.

Structure of the ERG gene

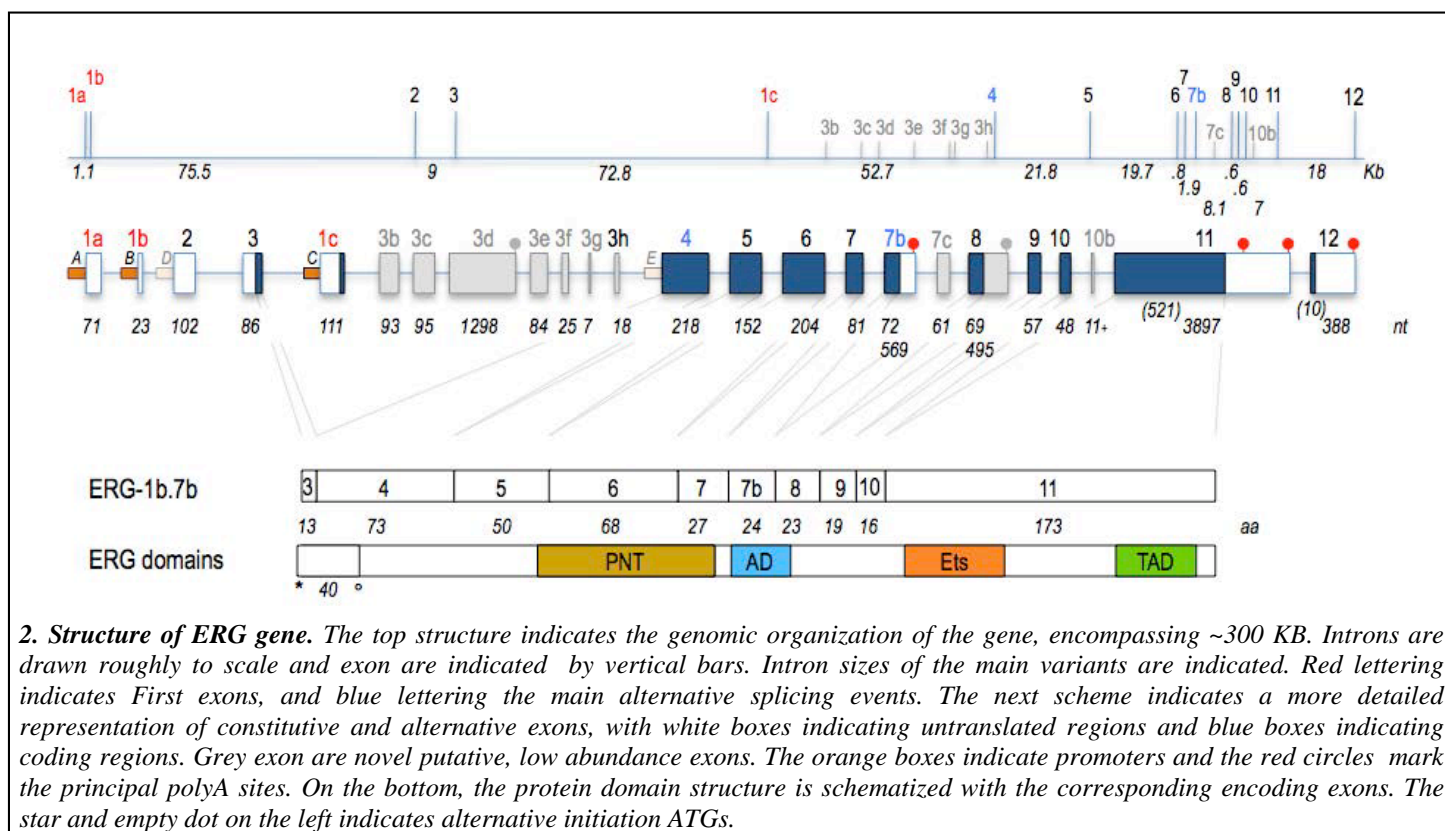
Multiple ERG isoforms can arise from the human ERG gene due to a combination of alternative transcription initiation sites, splicing and polyadenylation. These isoforms can in turn combine with TMPRSS2 and other 5' partners to produce a large number of ERG-derived variant mRNAs, with variable prognostic values [9, 10, 22]. In order to better understand the activities of various ERG-derived oncogenic products, we sought to initially clarify the ERG gene structure and to characterize the expression patterns of the variants expressed.

This is very complex, and one problem consists in the considerable contradictions in the ERG nomenclature, particularly regarding the identity of specific isoforms and their exons, with at least four different classification schemes. For example the large terminal exon that contains the Ets and transactivation domains is variably referred to as exon 11[2, 25], 12 [4], 16 [19, 26], 17 [21]. Similar discrepancies are also found when mRNAs or protein isoforms are involved, with the obvious consequence of generating confusion when trying to interpret the results of different groups which might adopt different conventions.

We worked out and describe in Figure 2 an up-to-date view of the exon-intron structure of the ~300 KB ERG locus (ENSG000000157554) with a proposed nomenclature that aims at incorporating the most established conventions based on prevalent literature, in order to minimize confusion.

The structure is mainly based on the 11 exons of the ERG Refseq entry NM_004449.4 (corresponding to Uniprot entry P11308 for ERG2), and incorporates the exon numbering used in the seminal Tomlins paper first describing the TMPRSS2:ERG fusion. In short, we maintain as exon 4 the 218 nt exon that is the main partner of TMPRSS2 exon 1 in the TMPRSS2:ERG fusions; we maintain as exon 11 the large 3897 nt last exon that encodes for the entire Ets domain; we named Exon 1a, 1b, 1c the three mutually exclusive 'first' exons following the three validated promoters PA, PB and PC (of note, Exon 1c is contained within intron 3 and splices directly to exon 4); alternative exons that are not part of the 11 reference exons are distinguished by letter variants so, for example, the 72nt alternative exon between exon 7 and 8 is hereby referred to as exon 7b and so on.

We then identified the main alternative regulatory events that generate the majority of ERG variability: 3 alternative promoters (PA, PB and PC); two common alternative splicing events (inclusion/skipping of exon 4 and 7b); three separate polyadenylation sites (7bpA, 11LpA and 12pA).

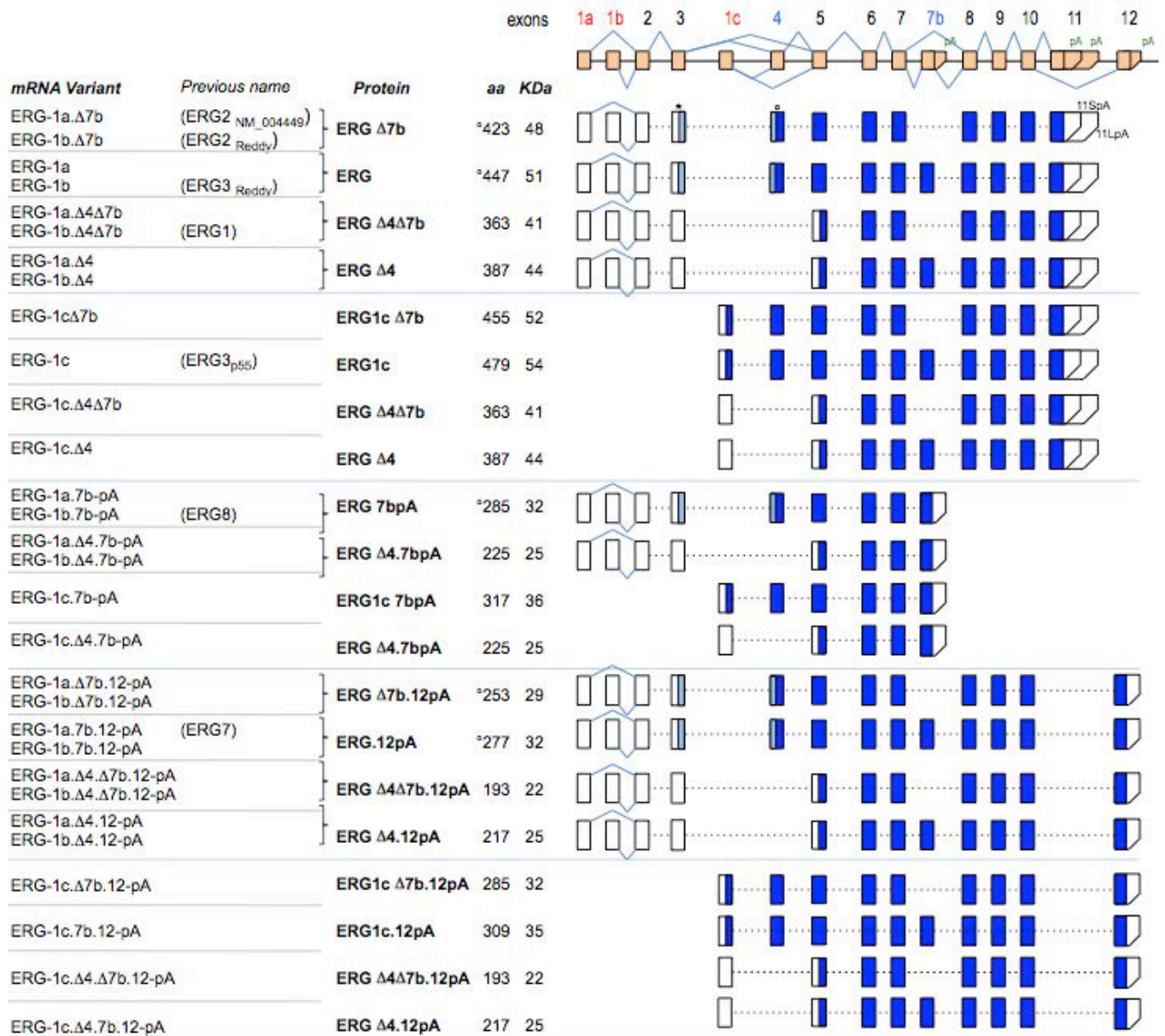


Combinatorial usage of the compatible alternative events generates 30 possible ‘major’ ERG transcript variants (Fig 3), which can encode 15 different predicted ERG-related polypeptides, with 3 different N-terminals, 3 different C-terminals and 1 possible internal variation (inclusion or skipping of 24 aa in the Alternative Domain encoded by Exon 7b).

In addition to the 30 ‘major’ variants described in figure 1, a plethora of ‘minor’ isoforms have been reported in literature or are present in databases (Suppl Fig 1). The list includes variants showing skipping of exons 2, 5, 7 and 8; usage of a proximal (Short) polyA site in exon 11 (11SpA) or of an additional intronic one downstream of exon 8 (8pA); and inclusion of supplementary alternative exons 7c, 10b and of multiple alternative exons in intron 3 (exons 3b-h), indicated in gray in figure 1. The evidence for the specific size of some of these exons is non conclusive as they seem to derive from partial cDNAs that start or end within the exons themselves. ERG 4, 6 and 9 from the previous Owczarek study have been reclassified in this group, whereas ERG 5 appears to be a truncated cDNA derived from one of the main isoforms.

Because any of these additional minor events could in principle combine independently with all the structurally compatible major isoforms, hundreds of variants could potentially be generated. However, all these events appear to occur very sporadically, and their physiological relevance is so far unknown.

The structure of the ERG gene was much more complex than originally anticipated, and its full elucidation reported here was a major undertaking that took significant amounts of time and resources.



3. Main ERG variants. ERG three promoters (pA, pB, pC), two main alternative splicing events (skipping/inclusion of exons 4 and 7b) and three main polyA sites (7bpA, 11LpA, 12pA) combine to generate 30 principal mRNA variants, which can encode 15 different ERG protein isoforms. The new proposed nomenclature, together with the previous known of the few described variants, the protein encoded, its size and predicted MW are indicated, along with a cartoon of the exon structure. Usage of exon 1a or 1b is always mutually exclusive and depends on the promoter engaged, and it always splices to exon 2, as indicated by the light blue chevrons.

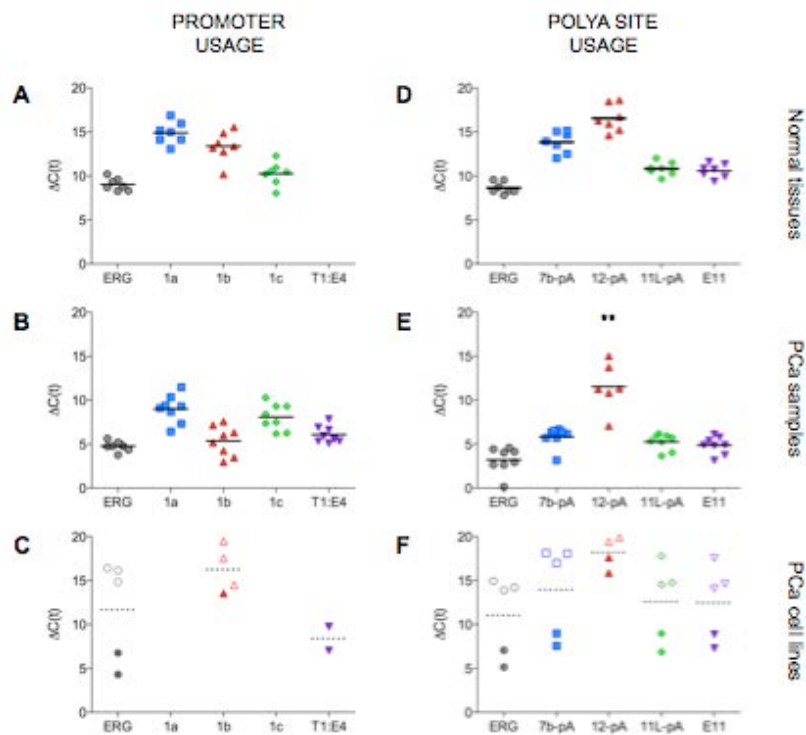
We have then proceeded to carry on characterization of only what appear to be the main variants, which are more likely to be of physiological significance in normal ERG functions and in disease

Characterization of ERG expression

Different tissues, tumors and cell lines were analyzed by qPCR to assess differences in usage of promoters and polyadenylation signals. While some degree of tissue to tissue variability is observed

(SupplFig 2), in general in normal tissues promoter PC (mean $\Delta C(t)$ = 10.2) appears to be the most active, being ~25-fold and ~10-fold more active than promoters PA ($\Delta C(t)$ = 14.9) and PB ($\Delta C(t)$ = 13.4), respectively (Figure 4A, lower point means higher expression levels). On the other hand, in a panel of 8 primary PCa samples expressing TMPRSS2-ERG fusions, promoter PB (mean $\Delta C(t)$ =5.4) accounts for the majority of native ERG transcript and it is present at levels comparable to those of the fusion itself ($\Delta C(t)$ =6), over 100-fold more abundant than the same transcript in normal prostate ($\Delta C(t)$ = 12.8) (Figure 4B and SupplFig 2A). Indeed, the native PB promoter appears to be the principal source of ERG transcript in at least some of the TMPRSS2:ERG carrying samples (SupplFig 2). Promoters PA ($\Delta C(t)$ =9) and to a lesser extent PC ($\Delta C(t)$ =8) are also activated in tumors when compared to normal prostate, but not as much as PB. In PCa cell lines, the preferential activation of PB is even more pronounced, as it is the only native ERG promoter active, while signals from PA and PC are undetectable (Figure 4C).

Two of the prostate cancer cell lines tested, VCap and NCI-H660 carry the TMPRSS2:ERG fusion. Accordingly, the fusion transcripts are very abundant (full symbols) in their RNA, while they are absent in that from the other PCa cell lines that do not carry the fusion (LNCap, DU145, C4-2,



4. Expression of ERG isoforms. (A-C) Expression levels of total ERG, and specific isoforms generated from promoters pA, pB and pC were measured by qPCR with specific primers, along with the expression of the fusion T1:T4 variant, in normal tissues (A), prostate tumor samples (B) and prostate cancer cell lines (C). (D-F) Expression levels of total ERG, and specific isoforms generated from polyA sites 7bpA, 12pA, 11LpA were measured by qPCR with specific primers, along with the expression of the exon 11, in normal tissues (A), prostate tumor samples (B) and prostate cancer cell lines (C). The indicated values in the graphs represent averages of at least 3 independent experiments and are presented as $\Delta C(t)$ normalized to GAPDH housekeeping gene, therefore a “high” $\Delta C(t)$ value means low levels of expression and a “low” value means high level of expression. See supplemental figure 2 and 3 for additional details.

open symbols and SupplFig 2). No expression from any of the native promoters was observed in the NCI-H660 cell line, which carries the fusion on both alleles, with loss of all the endogenous promoters [27].

Of the three principal polyadenylation signals described (Figure 2), the exon 11 Long polyA site (11L-pA) is needed to generate a fully functional ERG protein. The other polyadenylation sites in intron 7b (7b-pA) and exon 12 (12-pA) give rise to C-terminally truncated ERG isoforms lacking the functional Ets domain either because the transcripts terminate early (7b-pA) or because exon 11 is skipped when exon 12 is used (12-pA). In normal tissues, the 11L-pA signal generating the full-length transcript, was the most commonly used (mean $\Delta C(t)=10.8$), about 8-fold more than 7b-pA ($\Delta C(t)=13.8$) and 50-fold more than 12-pA ($\Delta C(t)=16.6$) (Fig 4D). Interestingly, in the TMPRSS2:ERG-expressing prostate tumors, the usage of 7b-pA, results strongly activated and is about as common as 11L-pA (Figure 4E). The same is true in PCa cell lines expressing the fusion (Figure 4F) suggesting that transcripts from the TMPRSS2:ERG fusion preferentially use the 7b-pA site, and that switching to this pattern of expression could be associated to tumor progression. A second, proximal polyA site producing a shorter 3' UTR had been described for exon 11 (11S-pA)[16, 19]. We indirectly assessed its use in our samples by quantitatively amplifying regions in exon 11 located before and after the proximal site (E11 and 11L-pA). Because the levels of expression in these two regions are the same in all samples analyzed, we conclude that usage of 11S-pA is marginal under most conditions (Fig 4D-F and SupplFig 3).

To determine if any of the alternative splicing events described above was differentially regulated in PCa, we analysed regions around exon 7b and exon 4 in normal tissues, PCa samples and cell lines by semi quantitative-PCR. Both alternative events were readily detectable (Figure 5), with a clear prevalence of the exon 4 and exon7b inclusion variants, but we didn't observe any significant

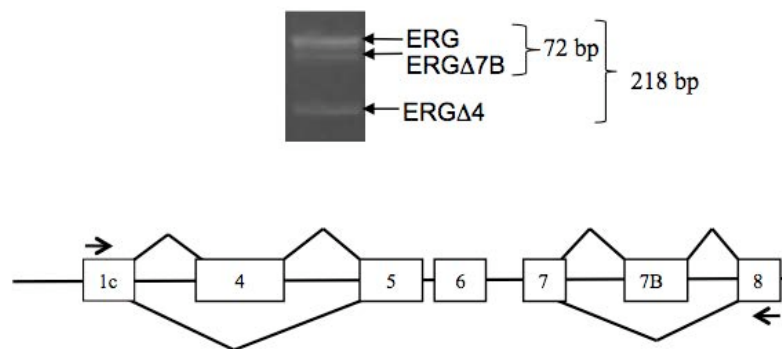


Figure 5. Alternative Splicing hot spots in ERG gene. PCR reaction spanning exons 1c to 8, or exons 3 to 8 reveal the consistent presence of multiple bands. Sequencing of these commonly observed alternatively spliced PCR products in multiple cell lines and tissues has led to the identification of inclusion/skipping of exon 4 (218 bp) and of exon 7b (72 bp) as splicing “hot spots” in the ERG gene, in both the normal and the fusion contexts. An example of a representative PCR reaction from RNAs derived from VCap cells using primers amplifying from exons 1c and 8 is shown (depicted by arrows).

difference in the relative amount of splicing variants between normal tissue and prostate samples, suggesting that the regulation of these two specific splicing events does not play a significant role in prostate cancer (data not shown).

Mapping of ERG and Tmprss2 –ERG starting ATG.

To evaluate whether the heterogeneity in the 5' region affected ERG expression and activity, we subcloned into a mammalian expression vector the native variants ERG-1a, ERG-1b, ERG-1c, ERG-1b.D4, ERG-1c.D4, and the common fusion variant T1:E4, with their complete 5' UTRs (Fig 6). In all cases exon 7b was included in the cDNA, as this is the most frequent configuration. Upon transient transfection in HeLa cells, expression of ERG-1c was efficient and corresponded to a peptide of the expected 54 KDa size (Fig. 7A). On the contrary, expression of the native ERG-1a and ERG-1b variants was inefficient, and resulted in peptides migrating at a size smaller than the 55 KDa expected if the first in-frame predicted ATG in exon 3 was used (Fig 7A, lanes 2-3). Multiple explanations could account for the unexpected gel mobility, such as the N-terminal conformation affecting migration, differential processing of the unique N-terminal or use of a downstream ATG. Interestingly, this peptide co-migrated with that generated by the transient over-expression of the T1:E4 fusion variant, at around 50 KDa (lane 5). Additional variation in the N-terminal region derives from exon 4 skipping which alters ERG's main ORF, so the predicted starting ATGs in exon 1c or

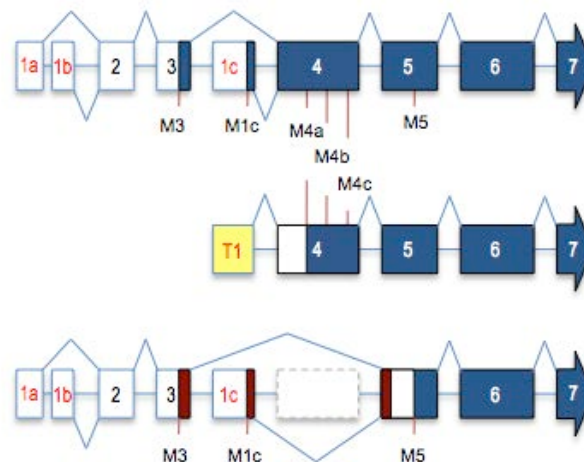


Figure 6. Alternative Splicing hot spots in ERG gene. PCR reaction spanning exons 1c to 8, or exons 3 to 8 reveal the consistent presence of multiple bands. Sequencing of these commonly observed alternatively spliced PCR products in multiple cell lines and tissues has led to the identification of inclusion/skipping of exon 4 (218 bp) and of exon 7b (72 bp) as splicing “hot spots” in the ERG gene, in both the normal and the fusion contexts. An example of a representative PCR reaction from RNAs derived from VCap cells using primers amplifying from exons 1c and 8 is shown (depicted by arrows).

exon 3 cannot generate an ERG-related peptide (Fig 6). Transient overexpression of ERG-1b.D4, ERG-1c.D4 results in both cases in peptides migrating at around 44KDa, consistent with the predicted usage of an in-frame ATG in exon 5 (Figure 7A, lanes 6-7). This is most evident when using EG antibody C-17 (Cell Signaling), which readily recognize recombinant ERG, but gives a very low signal of endogenous proteins from lysates from multiple cell lines. A different ERG antibody (anti-ERG C-20, Cell Signaling) preferentially recognizes in lysates an endogenous band corresponding in size to ERGD4 (lanes 1-5), but the precise identity of this product is unclear) so the switch to ATG M5 usage would be hard to detect.

Since T1:E4 lacks the predicted initiation codons from both TMPRSS2 and ERG transcripts, an alternative internal ATG from within ERG's open reading frame must be used to express an ERG-related product from the fusion mRNA, likely from exon 4. To identify the T1:E4 initiation codon, a series of methionine to alanine point mutations were generated in exon 4 of ERG for the three in-frame ATG codons encountered (Fig 6). Mutation at nucleotides 79 (M4a), but not 121 (M4b) and 184 (M4c) of exon 4 abolished T1:E4 expression (Figure 7B), indicating that the fusion transcript uses the

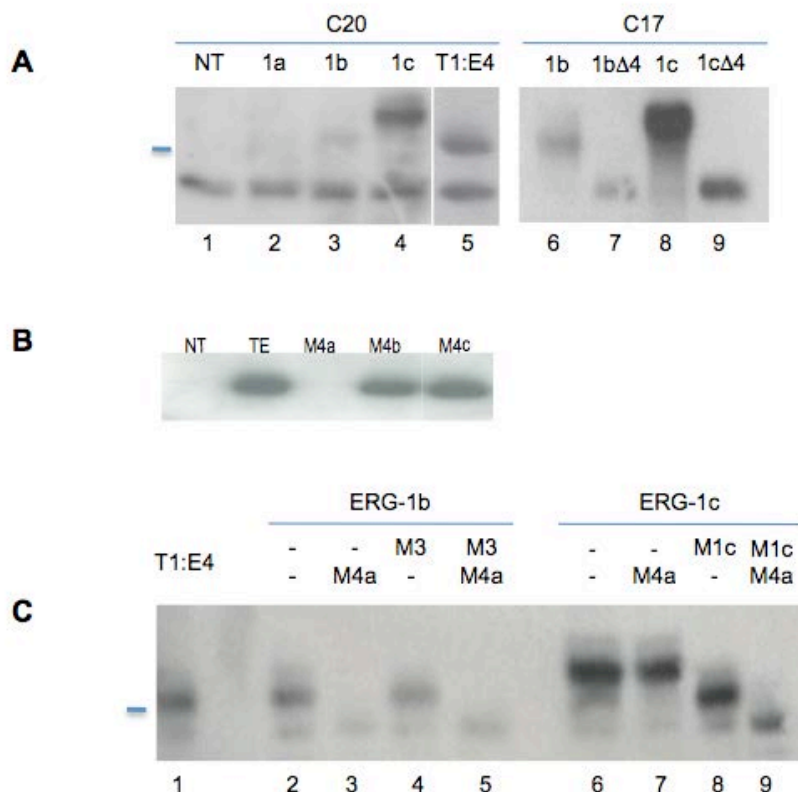


Figure 7. Alternative Splicing hot spots in ERG gene. PCR reaction spanning exons 1c to 8, or exons 3 to 8 reveal the consistent presence of multiple bands. Sequencing of these commonly observed alternatively spliced PCR products in multiple cell lines and tissues has led to the identification of inclusion/skipping of exon 4 (218 bp) and of exon 7b (72 bp) as splicing “hot spots” in the ERG gene, in both the normal and the fusion contexts. An example of a representative PCR reaction from RNAs derived from VCap cells using primers amplifying from exons 1c and 8 is shown (depicted by arrows).

first in-frame ATG for expression of the ERG peptide

Similar mutations in the first in-frame ATGs in exon 3, 1c and 4 were also introduced alone or in combination to map the starting ATGs from the native ERG isoforms (fig. 7C). Mutation of the ATG in exon 3 (M3) didn't have any effect on the expression of the 51 KDa product (Figure 7C, lanes 2 and 4), while mutation of the next ATG in exon 4 (M4a) abrogated it both in the wt and M3 context (Figure 7C, lanes 3 and 5), indicating that ERG-1b (and ERG-1a) do not use their first in-frame ATG, as would be predicted by the 5' cap dependent scanning model of translation initiation site selection[30], but instead it selects the following ATG in exon 4, generating a peptide identical to that encoded by the T1:E4 fusion. On the other hand, when the initial start codon in exon 1c is mutated (M1c), ERG-1c mobility is reduced from ~54 KDa to ~51 KDa (Figure 7C, lanes 6 and 8), like T1:E4 and ERG-1b (lane 1 and 2), consistent with a switch to the M4a ATG. Indeed, when this is mutated in the context of M1c (M1c/M4a) the 51 KDa product is also abrogated in favour of a smaller product, which appears to correspond to relatively inefficient usage of the ATG in exon 5. Altogether, these experiments show that native variants ERG-1c, ERG-1b.D4, ERG-1c.D4, as well as the common T1:E4 fusion, mainly use the first predicted in-frame ATG available, as expected from the scanning model. On the contrary, the ERG-1a and ERG-1b variants preferentially use the second in-frame ATG, and do so at a significantly lower efficiency.

Inhibition TMPRSS2:ERG initiation codon

One method to moderate the translation of the fusion-specific ERG isoform is through the use of morpholinos. Morpholinos are antisense-based compounds known to modulate alternative splicing through base pairing with their target RNA sequences and blocking the access of other molecules to this target site. The morpholino methodology can also be easily adapted to modulate translation initiation. It has been described that binding of morpholinos to the 5' untranslated region of the messenger RNA can block translation by inhibiting the progression of the ribosomal complex from the 5' cap to the start codon. Morpholinos that target the initiation ATG can also interfere with recognition of this specific site by the ribosome. However, the binding of

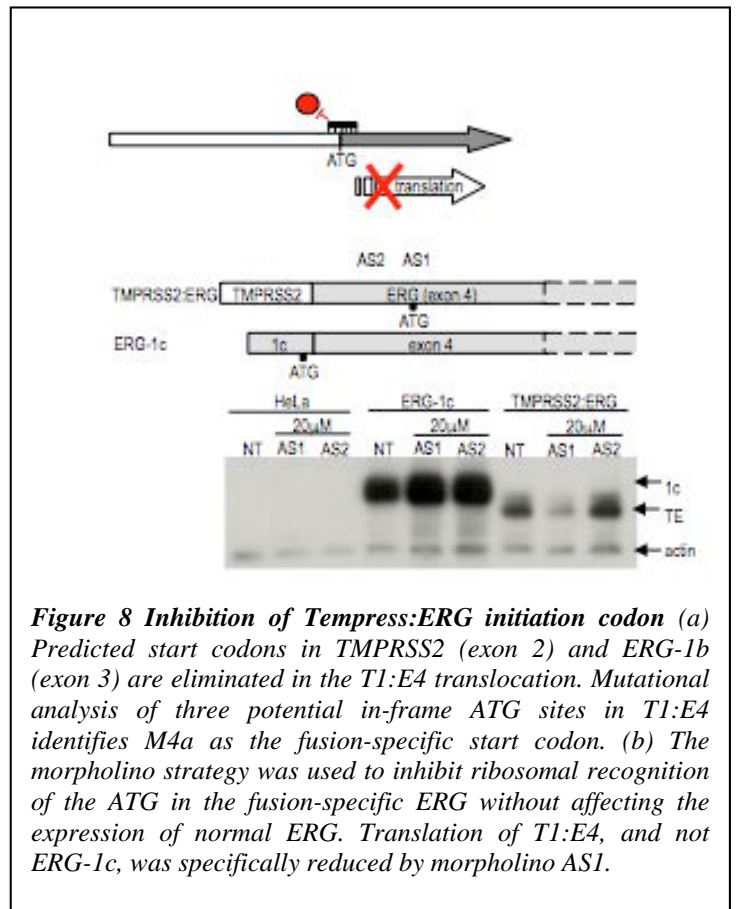


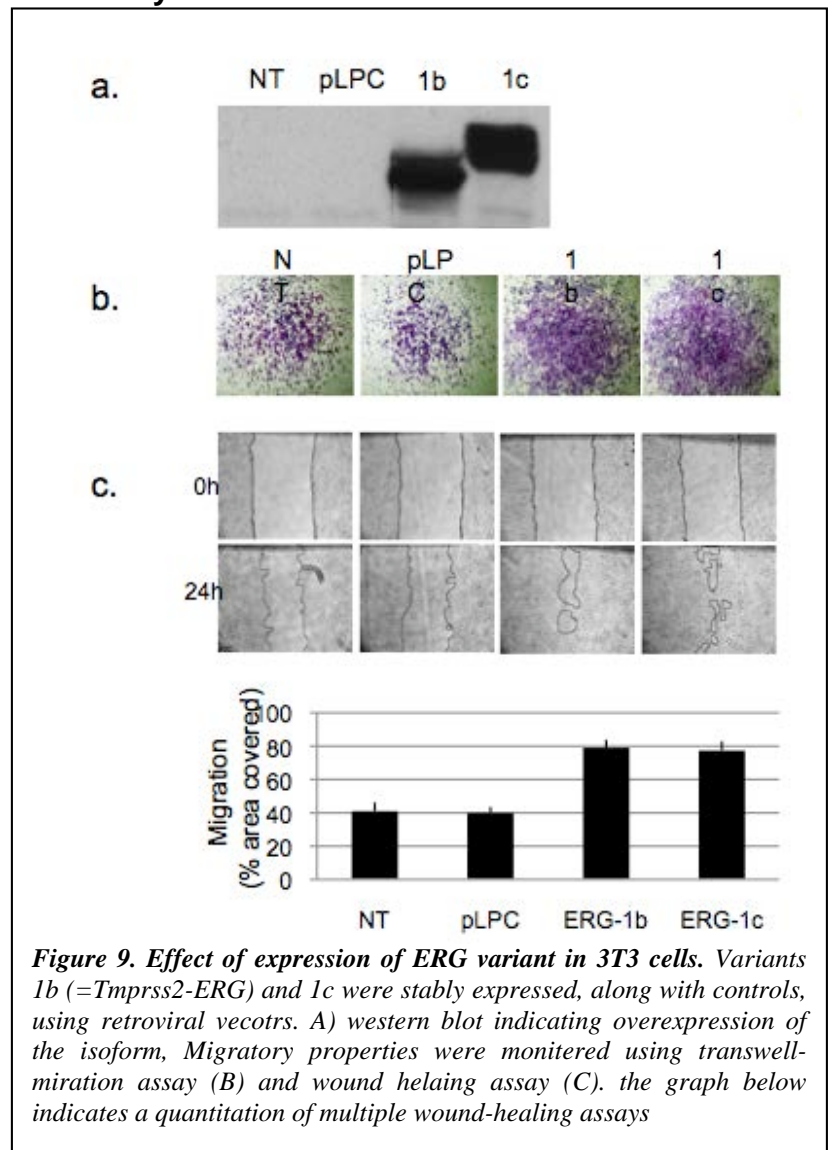
Figure 8 Inhibition of Tempress:ERG initiation codon (a) Predicted start codons in TMPRSS2 (exon 2) and ERG-1b (exon 3) are eliminated in the T1:E4 translocation. Mutational analysis of three potential in-frame ATG sites in T1:E4 identifies M4a as the fusion-specific start codon. (b) The morpholino strategy was used to inhibit ribosomal recognition of the ATG in the fusion-specific ERG without affecting the expression of normal ERG. Translation of T1:E4, and not ERG-1c, was specifically reduced by morpholino AS1.

morpholinos to a sequence downstream of the initiation codon would be displaced by the incoming mature ribosome and translation of the protein would not be affected. In order to selectively block TMPRSS2:ERG translation, we designed two morpholinos: AS1, which targets the start codon in the fusion context (M4a), and AS2, which targets a sequence just upstream of the initiation ATG (Figure 8b). AS1 binds to the T1:E4 initiation codon and was able to efficiently block recognition of this ATG by the ribosome and inhibit translation of the fusion-specific ERG. The translation of ERG-1c was unaffected since the mature ribosome forms in exon 1c and is able to displace the AS1 morpholino located in exon 4 downstream. Morpholino AS2 which targets a sequence 24 base pairs upstream of M4a had no effect on the translation of ERG-1c. Translation of T1:E4 however, was not blocked by AS2. Although morpholinos that bind anywhere within the 5' untranslated region are likely to prevent formation of the mature initiation ribosome, it is likely that double-stranded regions formed in the secondary structures of the T1:E4 messenger RNA are less accessible by the morpholino.

TMPRSS2:ERG fusion variants biological activity

To assess whether the described structural differences at the N-terminal affect ERG biological activity, the ERG cDNAs described above were subcloned into a retroviral vector and stable overexpressing clones were generated in NIH-3T3 cells (figure 9A). In agreement with previous reports showing that the TMPRSS2:ERG fusion affects cell motility, we also observed promotion of in vitro cell invasion and migration, as assayed by transwell migration through matrigel and scratch wound assay (Fig 9B-C), by all the tested ERG variants, regardless of the identity of the N-terminal domain.

In addition, overexpression of the isoforms also slowed down cell proliferation (Fig 10A), but was associated with a reduction in apoptosis (Fig 9B), consistent with an observed increase in cellular senescence (Fig 10C). None of the



isoforms, however, was sufficient to mediate cellular transformation and promote anchorage-independent growth in soft-agar (not shown).

Although the characterization of the ERG isoforms at the protein level enabled us to better understand the regulatory mechanisms underlying their expression, the primary source of ERG in prostate cancer is through the formation of the androgen-driven TMPRSS2:ERG fusions. Because some fusion isoforms are described to be more highly expressed, as well as some isoforms being associated with a more aggressive cancer phenotype, it was important to assess the various fusions isoforms isolated for any biological function.

We thus extended the previous experiments to a broader panel the various TMPRSS2:ERG fusion transcripts, to evaluate their biological activity and assess their functional significance in prostate carcinogenesis. All the main isoforms, indicated in Figure 11, were stably expressed and their invasion/migratory properties were assayed as above (Figure 11 A-C).

Furthermore, to directly measure the transcriptional activities, we subcloned the VE-cadherin and the matrix metalloproteinase-1 (MMP-1) promoters which are two very well-characterized downstream targets of ERG (24-26) (Figure 10D). Wound healing and a transcription dual-luciferase assays were used to assess the biological activity of the various fusion isoforms identified. The rates of wound healing by the specific fusion isoforms correlated fairly well with the luciferase activity generated. Although some isoforms have been described to be more tumorigenic than others, and some isoforms are more highly represented in distinct prostate cancers, it is difficult to speculate on the importance of the activity generated by one specific isoforms relative to another.

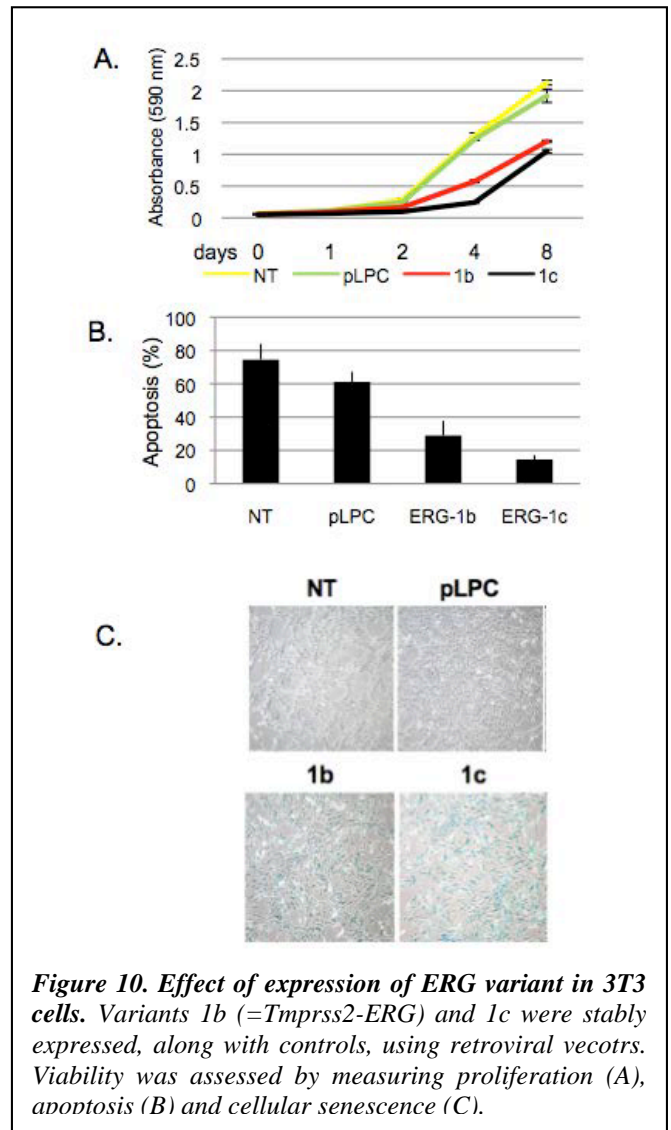


Figure 10. Effect of expression of ERG variant in 3T3 cells. Variants 1b (=Tmprss2-ERG) and 1c were stably expressed, along with controls, using retroviral vecotrs. Viability was assessed by measuring proliferation (A), apoptosis (B) and cellular senescence (C).

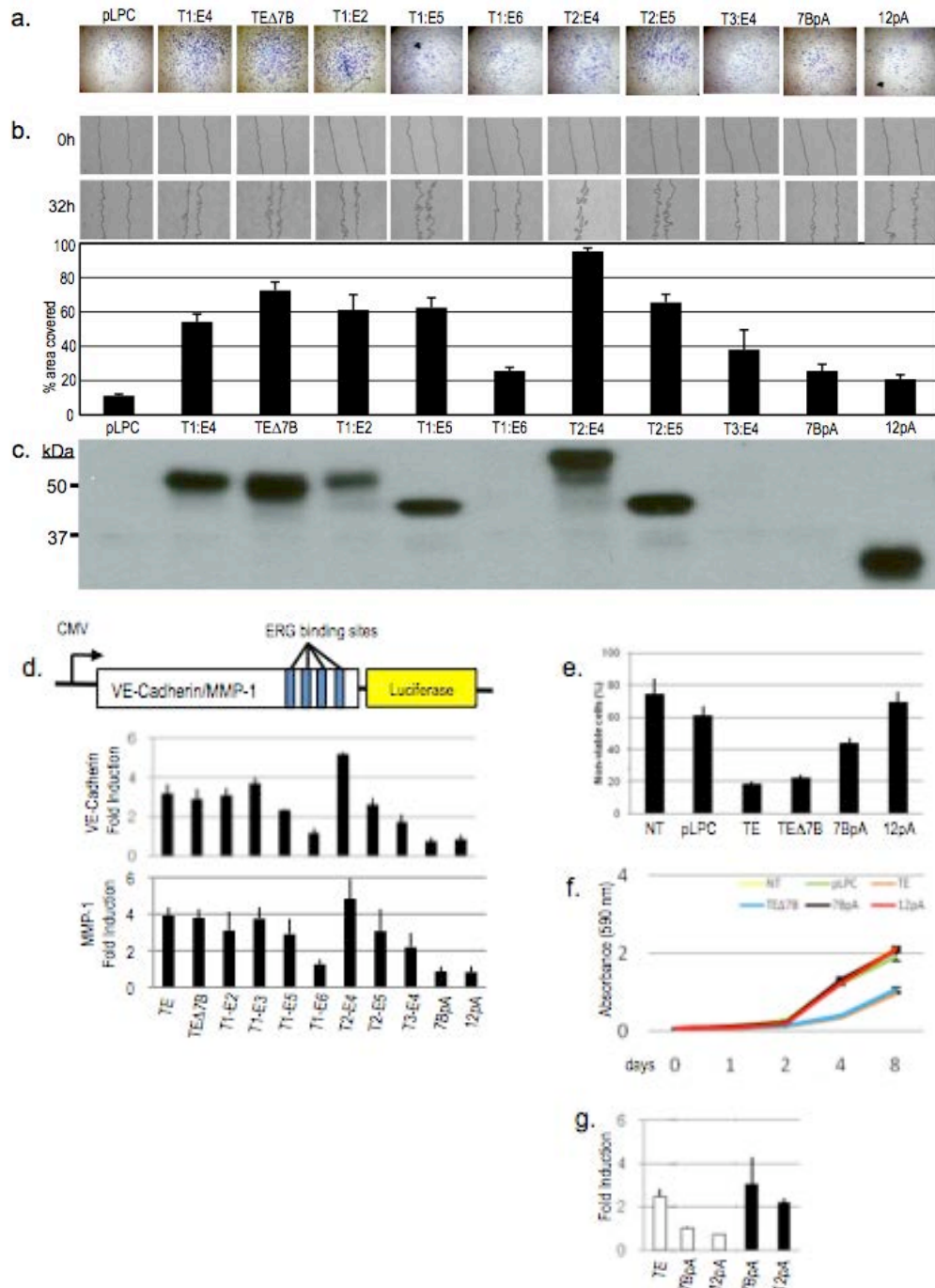


Figure 11. Biological activity of fusion isoforms. The cellular invasion (a) and migration rates (b-c) of the fusion isoforms was assessed in stable NIH-3T3 cells. Isoforms which did not express a peptide, or that lacked the Ets/TAD domains displayed rates similar to the vector control (T1:E6, 7BpA and 12pA). The remaining peptides displayed higher rates of migration and invasion depending on the level of protein expression (c). The activity of the fusion isoforms was further assessed through the activation of luciferase reporters containing ERG target sites, MMP-1 and VE-cadherin (d). The luciferase activity generated is also dependent on the expression levels of the fusion isoforms, with activity levels similar to the rates of migration. (e) Trypan blue cell death and (f) cell proliferation assays were performed to evaluate a potential dominant negative role for 7BpA and 12pA. The truncated isoforms did not protect cells from dying and did not exert any effects on the rates of cell proliferation. (g) Luciferase activity of fusion isoforms alone (white bars) or co-expressed with TE (black bars). Co-expression of either 7BpA or 12pA with TE did not inhibit the luciferase activity generated by TE.

The fusion isoforms for which translation initiates in exon 4 (T1:E4, TEΔ7B, T1:E2 and T1:E3) or in exon 5 (T1:E5 and T2:E5) displayed similar rates of wound closure (Fig.10a, lanes 2-6 and 9). Similarly, the luciferase activity generated was comparable, except in the context of the VE-cadherin

promoter, which shows lower luciferase activity for T1:E5 and T2:E5 (Fig.11 d lanes 1-5 and 8). The moderate levels of luciferase activity of T3:E4, as well as 40% higher migration rate relative to the vector control, confirms that protein translation does occur and generates a functional peptide.

Similarly, T1:E6 for which protein detection was not possible, generated very low migration levels and luciferase activity, but still greater than the vector control. The two isoforms in which the Ets and TAD functional domains are lacking, TE:7BpA and TE:12pA, did not exhibit any significant rate of migration or luciferase activity. The truncated fusion isoforms further displayed cell proliferation and cell viability measures similar to the vector control, relative to T1:E4 and TE Δ 7B in which the functional domains are expressed (Fig.10 e-f). Since we suggested that TE:7BpA and TE:12pA might have dominant-negative properties, we sought to assess this characteristic through the inhibition of T1:E4 luciferase activity. Co-expression of TE:7BpA or TE:12pA with T1:E4 generated significant luciferase activity, which was solely derived from the expression of T1:E4 without any inhibitory contribution from either of the truncated isoforms (Figure 11g). This would suggest that TE:7BpA and TE:12pA produce inactive proteins, with no biological contribution.

KEY RESEARCH ACCOMPLISHMENTS

- redefinition of the genomic structure of the ERG gene
- characterization of ERG variants expression in normal tissues and prostate cancers
- identification of endogenous ERG as a major source of Erg expression, in addition to Tmprss2:Erg
- identification of a truncated variant (7b-pA) as one of the most abundant isoforms in tumors
- Mapping of the ATG starting sites in various variants, including the oncogenic ones
- Specific inhibition of the oncogenic ATG in cells
- Characterization of biological activity of various isoforms in vitro

REPORTABLE OUTCOMES

Abstract/presentations

Boutsalis, et al. **Oncogenic potential of multiple splicing isoforms of the prostate-cancer specific *TMPRSS2-ERG* gene fusion products** RNA Society's 2010 Annual Meeting, Seattle, (2010).

A manuscript is in preparation and planned for submission before the summer.

CONCLUSIONS

Although several groups have tried to correlate the expression of the TMPRSS2:ERG fusion to the clinical status of prostate cancer, or to characterize it as a prognostic marker, the results have resulted in confounding outcomes. This suggests that the involvement of TMPRSS2:ERG in prostate cancer is complex and further studies would be required to elucidate its prognostic value, if any. For instance, the deletion of the genome between TMPRSS2 and ERG has been associated with a poor clinical outcome, in the same way as a poor prognosis has also been associated with the presence of specific TMPRSS2:ERG isoforms (27,28).

Overexpression of ERG in prostate cancer had been previously described [26], but since the discovery of the TMPRSS2:ERG fusion, the increased levels in ERG transcripts have been ascribed entirely to this event, especially since the approaches used typically don't allow distinction between native and fusion transcripts. However, our novel observation that in addition to the fusion-derived transcripts, native ERG can also be highly overexpressed in PCa is very important as it might in part redefine the role that native ERG plays in prostate cancer development. Indeed, a key role for native ERG to accelerate initiation and promote progression of prostatic adenocarcinomas is underscored by the observation that overexpression of the endogenous mouse *Erg* transcripts is significantly increased in tumors from prostate conditional *Pten*^{-/-};*Trp53*^{-/-} mice, compared to *Pten*^{-/-};*Trp53*^{+/+} mice [8].

While the latter result in an indolent form of PCa, the former –and the associated *Erg* overexpression- result in an aggressive, lethal phenotype [8]. The differential activation of the three native promoters in the fusion-carrying PCa samples also suggests that this aberrant expression is generated by regulatory events at the level of transcription rather than by genomic ones, such as amplification of the ERG locus. An intriguing possibility is that activation of the native ERG promoters, and of PB in particular, may be driven directly or indirectly by ERG itself as part of a positive regulatory loop. For example, several putative C-MYC responsive elements have been identified in the region immediately upstream of the ERG PB promoter [19]. Since the C-MYC oncogene has also been described as a key downstream target of ERG activity[28], the androgen-dependent activation of ERG from the fusion, or a separate PTEN-dependent MYC activation[29] could trigger a self-sustaining ERG/MYC oncogenic loop, which could eventually also become androgen-independent.

In addition, we have completed a detailed mapping of the ATG signals at the N-terminal of the ERG variants and the *tprss-erg* fusions, which identified the key ATG from the oncogenic product expression. We have confirmed and expanded our approach to specifically inhibit this ATG in vitro

and will continue to develop this approach and move it to in vivo context. We are also continuing the investigation on the role of the different 5'UTR on the isoforms expression levels and the characterization of the biological activities of the various variants by themselves or in combination.

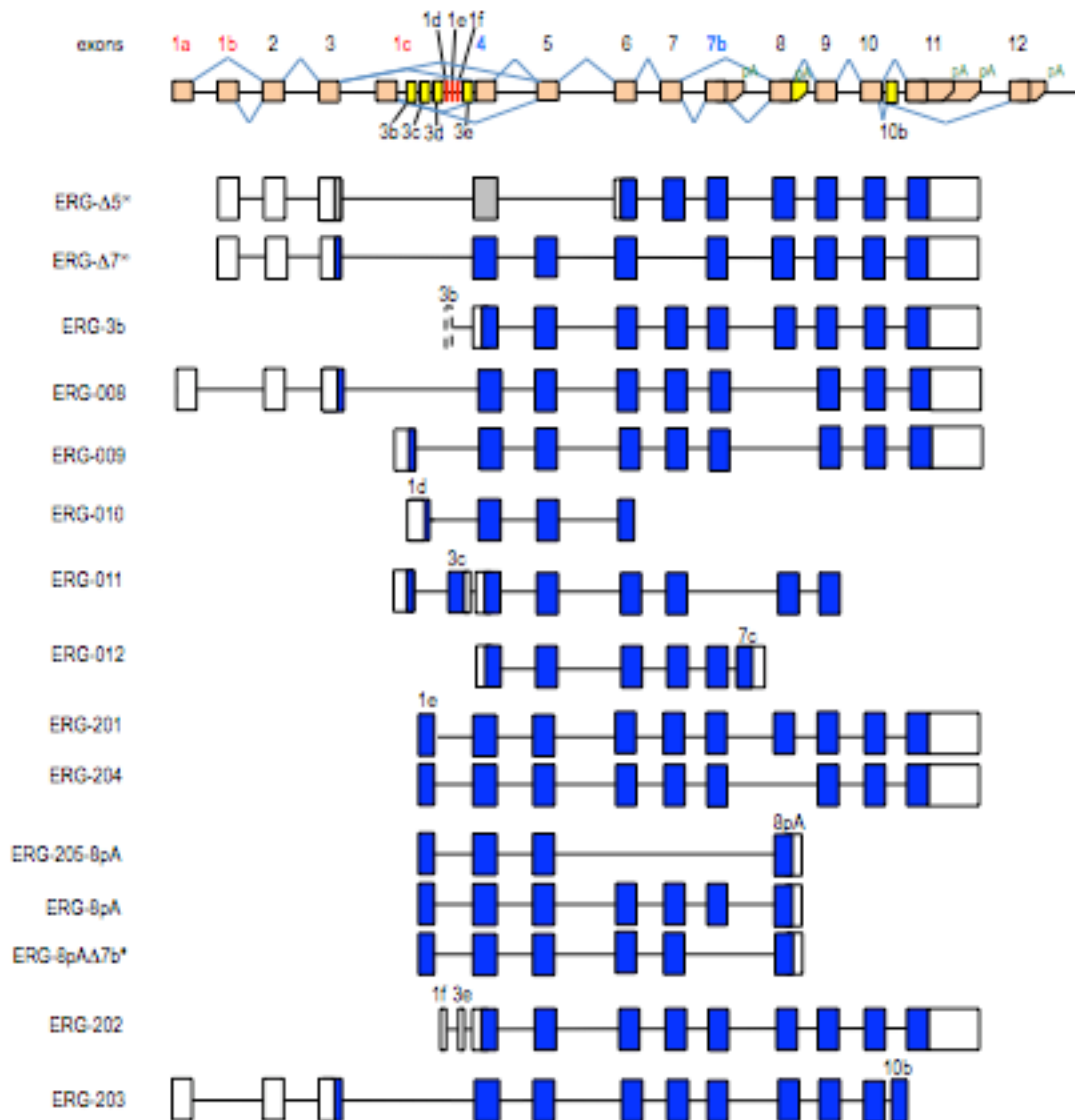
REFERENCES

1. Rowley, J.D., *Chromosome translocations: dangerous liaisons revisited*. Nat Rev Cancer, 2001. **1**(3): p. 245-50.
2. Tomlins, S.A., et al., *Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer*. Science, 2005. **310**(5748): p. 644-8.
3. Shand, R.L. and E.P. Gelmann, *Molecular biology of prostate-cancer pathogenesis*. Curr Opin Urol, 2006. **16**(3): p. 123-31.
4. Clark, J.P. and C.S. Cooper, *ETS gene fusions in prostate cancer*. Nat Rev Urol, 2009. **6**(8): p. 429-39.
5. Tomlins, S.A., et al., *TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer*. Cancer Res, 2006. **66**(7): p. 3396-400.
6. Tomlins, S.A., et al., *Role of the TMPRSS2-ERG gene fusion in prostate cancer*. Neoplasia, 2008. **10**(2): p. 177-88.
7. Helgeson, B.E., et al., *Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer*. Cancer Res, 2008. **68**(1): p. 73-80.
8. Carver, B.S., et al., *Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate*. Nat Genet, 2009. **41**(5): p. 619-24.
9. Demichelis, F., et al., *TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort*. Oncogene, 2007. **26**(31): p. 4596-9.
10. Nam, R.K., et al., *Expression of TMPRSS2:ERG gene fusion in prostate cancer cells is an important prognostic factor for cancer progression*. Cancer Biol Ther, 2007. **6**(1): p. 40-5.
11. Perner, S., et al., *TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer*. Cancer Res, 2006. **66**(17): p. 8337-41.
12. Rajput, A.B., et al., *Frequency of the TMPRSS2:ERG gene fusion is increased in moderate to poorly differentiated prostate cancers*. J Clin Pathol, 2007. **60**(11): p. 1238-43.
13. Klezovitch, O., et al., *A causal role for ERG in neoplastic transformation of prostate epithelium*. Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2105-10.
14. Trotman, L.C., et al., *Pten dose dictates cancer progression in the prostate*. PLoS Biol, 2003. **1**(3): p. E59.
15. Squire, J.A., *TMPRSS2-ERG and PTEN loss in prostate cancer*. Nat Genet, 2009. **41**(5): p. 509-10.
16. Rao, V.N., T.S. Papas, and E.S. Reddy, *erg, a human ets-related gene on chromosome 21: alternative splicing, polyadenylation, and translation*. Science, 1987. **237**(4815): p. 635-9.
17. Prasad, D.D., et al., *Differentially spliced erg-3 product functions as a transcriptional activator*. Oncogene, 1994. **9**(2): p. 669-73.
18. Duterque-Coquillaud, M., et al., *New human erg isoforms generated by alternative splicing are transcriptional activators*. Oncogene, 1993. **8**(7): p. 1865-73.
19. Owczarek, C.M., et al., *Detailed mapping of the ERG-ETS2 interval of human chromosome 21 and comparison with the region of conserved synteny on mouse chromosome 16*. Gene, 2004. **324**: p. 65-77.
20. Carrere, S., et al., *Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes via two distinct domains*. Oncogene, 1998. **16**(25): p. 3261-8.
21. Wang, J., et al., *Pleiotropic biological activities of alternatively spliced TMPRSS2/ERG fusion gene transcripts*. Cancer Res, 2008. **68**(20): p. 8516-24.
22. Wang, J., et al., *Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer*. Cancer Res, 2006. **66**(17): p. 8347-51.
23. Clark, J., et al., *Diversity of TMPRSS2-ERG fusion transcripts in the human prostate*. Oncogene, 2007. **26**(18): p. 2667-73.
24. FitzGerald, L.M., et al., *Association of TMPRSS2-ERG gene fusion with clinical characteristics and outcomes: results from a population-based study of prostate cancer*. BMC Cancer, 2008. **8**: p. 230.
25. Jhavar, S., et al., *Detection of TMPRSS2-ERG translocations in human prostate cancer by expression profiling using GeneChip Human Exon 1.0 ST arrays*. J Mol Diagn, 2008. **10**(1): p. 50-7.

26. Petrovics, G., et al., *Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome*. *Oncogene*, 2005. **24**(23): p. 3847-52.
27. Mertz, K.D., et al., *Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model*. *Neoplasia*, 2007. **9**(3): p. 200-6.
28. Sun, C., et al., *TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation*. *Oncogene*, 2008. **27**(40): p. 5348-53.
29. Ghosh, A.K., et al., *PTEN transcriptionally modulates c-myc gene expression in human breast carcinoma cells and is involved in cell growth regulation*. *Gene*, 1999. **235**(1-2): p. 85-91.
30. Kozak, M., *Regulation of translation via mRNA structure in prokaryotes and eukaryotes*. *Gene*, 2005. **361**: p. 13-37.
31. Chen, Z., et al., *Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis*. *Nature*, 2005. **436**(7051): p. 725-30.
32. Kozak, M., *Structural features in eukaryotic mRNAs that modulate the initiation of translation*. *J Biol Chem*, 1991. **266**(30): p. 19867-70.
33. Kozak, M., *Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits*. *Cell*, 1980. **19**(1): p. 79-90.
34. Kochetov, A.V., et al., *uORFs, reinitiation and alternative translation start sites in human mRNAs*. *FEBS Lett*, 2008. **582**(9): p. 1293-7.
35. Calvo, S.E., D.J. Pagliarini, and V.K. Mootha, *Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans*. *Proc Natl Acad Sci U S A*, 2009. **106**(18): p. 7507-12.

APPENDICES

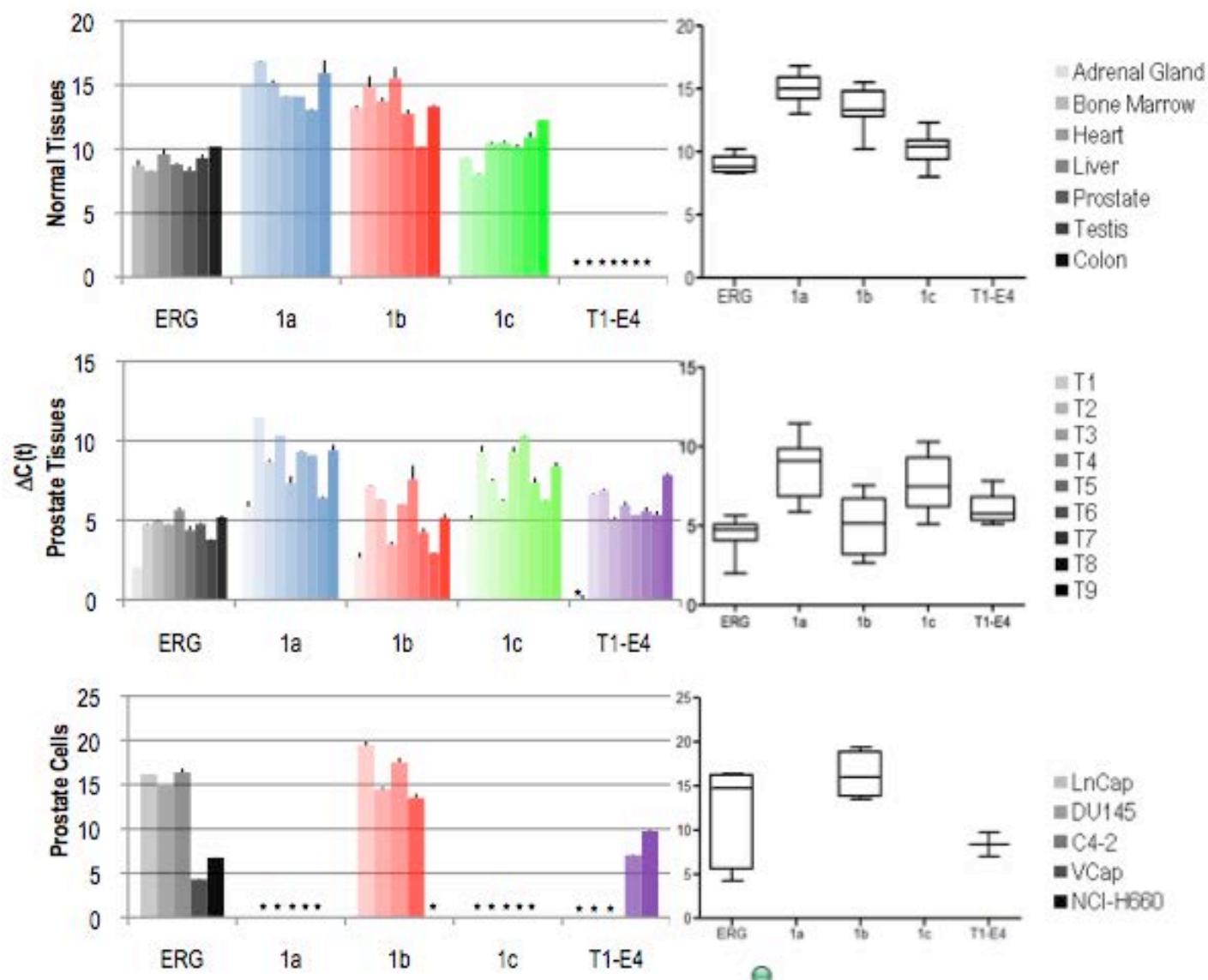
SUPPORTING DATA



Supplemental Figure 14. ‘Minor’ ERG Isoforms.

ERG exons and isoforms that have been described and listed in the Ensembl genome browser, or that we have identified (exons are shown in yellow and novel first exons are shown in red in the ERG locus schematic, top). The expression of these transcripts has yet to be proven (with the exception of the retention of intron 8 indicated in 8-pA). White boxes represent untranslated portions and those in blue represent the predicted translated proteins from the identification of in-frame ATG codons. Shaded region in ERG-Δ5 represents the introduction of a premature stop codon when exon 5 is skipped in the context of 1b.

(∞ The alternative splicing of exons 5 and 7 are shown in the context of 1b, but can also occur in the context of 1a and 1c. *Polyadenylation site 8-pA with the alternative splicing of exon 7B has been confirmed but not in the context of exon 1e.)



Supplemental Figure 2. Quantification of Alternative Promoter Use.

Promoter P_C is the most active in normal tissues, while promoter P_B is the most active in cancer tissue and the only one active in prostate cancer cell lines.

qPCR analysis of the alternative promoter use in a panel of normal tissues, prostate cancer samples and prostate cancer cells. Primer sets specific for the 3 different ERG promoters and for the TMPRSS2:ERG fusion where used, along with a primer set spanning exons 5-7 to quantify total ERG. The indicated values in the graphs on the left represent averages of 3 independent experiments and are presented as $\Delta C(t)$ normalized to GAPDH housekeeping gene, therefore a “high” $\Delta C(t)$ value means low levels of expression and a “low” value means high level of expression. Asterisks indicate that the gene was NOT detected in the sample and would therefore be equivalent to a column that goes to the top of the table, but it is omitted for clarity. The graphs on the right summarize the data on the left in a ‘box and whiskers’ representation with indicated the median and the 5th and 95th percentile value. NCI-H660 cells only express ERG from the TMPRSS2:ERG fusion because the fusion is present on both alleles and therefore the natural ERG promoters are completely absent.

